

TM-1276

多層記述を利用したタンパク質の
立体構造予測

鬼塚 健太郎

© Copyright 1993-07-07 ICOT, JAPAN ALL RIGHTS RESERVED

ICOT

Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan

(03)3456-3191~5

Institute for New Generation Computer Technology

多層記述を利用したタンパク質の立体構造予測

鬼塚健太郎

(財) 新世代コンピュータ技術開発機構

1 はじめに

タンパク質はそのアミノ酸配列によって機能、構造が決定される。したがって、配列が同じなら立体構造も同じことになり、立体構造は原理的にアミノ酸配列からほぼ一意に決定できることになる。つまり、アミノ酸配列と立体構造との対応がつかめれば、アミノ酸配列からの立体構造予測は可能なわけである。

今回は、これまで多くの研究者によって行われてきた局所アミノ酸配列からの局所構造予測方法とは大幅に異なるアプローチによって、タンパク質の立体構造をアミノ酸配列から直接決定する方法について述べる。

2 タンパク質の立体構造記述法

これまでの立体構造予測では、特徴的な局所構造である二次構造を配列から予測するという、いわゆる二次構造予測に主眼が置かれてきた ([Chou 74] や [Fasman 89])。しかし、この方法では、以下の二つの大きな問題がある。

- タンパク質の立体構造形成で重要な大局的な相互作用を考慮することができない [Branden と Tooze 91]。これは、アミノ酸配列上では遠い位置にあるアミノ酸同士が、3次元空間では曲がりくねることによって空間的に近い位置に存在しうることから、起こる相互作用である。
- 予測が正しくとも、予測結果から3次元の立体構造を構築できない。これは、二次構造が分かっても、そこから3次元構造を起こすには、べつのファクタを考慮した三次構造予測技術 [Cohen 等 82] が必要となることを意味している。

そこで、我々は、この2点を考慮して、まず、大局的な相互作用が考慮でき、さらに、元の立体構造を復元することが可能なタンパク質立体構造記述法 [Onizuka 等 93] [鬼塚 93] を開発した。これはタンパク質の立体構造を規模の異なる複数の階層で表現するものである。それぞれの層では、タンパク質の構造は記号列で表現され、それぞれの記号は、それぞれの層に固有な規模の局所構造の種類を表している。

まず、連続する5残基からなる局所構造を16種類に分類し、同じく、9, 17, 33, 65, 129残基からなる局所構造をそれぞれ16種類に分類する。次に、この分類された構造を表す記号の列によってそれぞれの規模の層でタンパク質の立体構造を表現するわけである。

以下は、比較的小さいタンパク質である、トリプシンインヒビタの立体構造をこの記述方法で記述したものである。

```
5  BAAGIHLFPOEJEIECFOFOPFDANEILOFPDGGMGJHIBGJCDJBAAAAAAGM
9  FCFLHDIPPNPLH000I1KMGLH0I0PPNNNNNEJ1JLFKMEBBAACF
17 IIGMMNLEFFHIGGJJJLEEFFDMMMNPPPO0000ABBPP0
33 NNNNGHHHHHKKOGGGGMMMF00
```

この記述は3次元構造の情報をかなり多く含んでいるので、これから元の立体構造を復元することは、ある程度の誤差で可能である。

3 統計的にモデル化された幾何学的制約

この新しいタンパク質立体構造記述法から3次元的に健全な立体構造を満たすべき幾何学的制約を取り出すことができる。たとえば、上の構造記述で、5残基層の最後の部分に近い所にAが並んでいるが、これはヘリックスである。これは連続的な構造であるから、Aの隣にはAがきやすい。また、5残基層での記述でAが幾つか並んだときは9残基層でもAになっている。これは9残基層ではヘリックスはAと分類されているからである。このことから、同じ層で近傍にきやすい構造はどのようなものであるか、あるいは、上下の層同士で同じ場所にくるのはどのような構造であるかを統計的に解析することで、タンパク質の立体構造の幾何学的制約をモデル化することができる。

4 幾何学的制約を考慮した立体構造予測

今回我々が開発した立体構造予測システムでは、確率モデルによって部分アミノ酸配列とその部分の構造との関係、すなわち一次制約をモデル化する。このモデル化された一次制約から、それぞれの部分の立体構造がどのようなものになりやすいかをもちめる。つまり、その部分配列がAという構造になる可能性、Bという構造になる可能性などを評価する。次に、幾何学的制約を考えて、ある部分構造がその部分構造と重なり合う全ての近傍の構造との間で幾何学的制約関係をみたしているかを調べ、全体として最もよく一次制約と幾何学的制約を満たす全体構造をもちめる。

計算アルゴリズムは、現在のところ、反復法を用いているが、線形計画法が有効である可能性もある。また、ニューラルネットなどを応用することも可能であると考えられる。

5 結語

現状では、予測精度はよくない。理由の一つとしては、確率モデルによって配列から構造の可能性を計算する部分が十分に良い性能をだしていないことによると思われる。また、本来復元性があるように作られた記述法であるにも関わらず、クラスタリングによる構造の分類に伴う誤差により、復元がかなり厄介である。今後の課題としては、復元性のある記述を徹底するためにも、記述の層の数を増やし、上下の位置関係がある程度の誤差で定まるようにすることがあり、また、一次制約のモデルの方法もいろいろなものを検討して見る必要がある。

参考文献

- [Cohen 等 82] Cohen, F.E., M.J.E. Sternberg, W.R. Taylor, "Analysis and prediction of the packing of α -helices against a β -sheet in the tertiary structure of globular proteins" in *J. Mol. Biol.* 156, 1982, pp. 821-862.
- [Fasman 89] Fasman, G.D.(editor), *Prediction of Protein Structure and the Principles of Protein Conformation*, 1989, New York: Plenum Publishing Corporation.
- [Branden と Tooze 91] Branden, C. and J. Tooze, *Introduction to Protein Structure*, 1991, New York: Garland Publishing, Inc.
- [Chou 74] Chou, P. Y. and Fasman, G.D. "Prediction of protein conformation" in *Biochemistry* 13, 1974, pp. 222-244.
- [鬼塚 93] 鬼塚健太郎, 「挑戦! 境界領域創造 - めざせ! タンパク質立体構造予測の新方法開発 -」, *bit* 25-6, 1993, pp 75-85.
- [Onizuka 等 93] Onizuka, K., K. Asai, M. Ishikawa and S.T.C. Wong, "A Multi-Level Description Scheme of Protein Conformation", in *Proc. of 1st Intl. Conf. of ISMB*, 1993.
- [Onizuka, Asai と Ishikawa 93] Onizuka, K., K. Asai, and M. Ishikawa, "A Scheme for Protein Tertiary Structure Prediction Based on Stochastic Reasoning", submitted to *WORKSHOP "ARTIFICIAL INTELLIGENCE and the GENOME" at IJCAI*, 1993.